Published in final edited form as:
IEEE EMBS Int Conf Biomed Health Inform. 2016 February ; 2016: 513–516. doi:10.1109/BHI.2016.7455947.

Informatics-based Challenges of Building Collaborative Healthcare Research and Analysis Networks from Rural Community Health Centers

Daniel R. Harris,
Center for Clinical and Translational Sciences, University of Kentucky, Lexington, Kentucky 40506

Tamela J. Harper,
Center for Clinical and Translational Sciences, University of Kentucky, Lexington, Kentucky 40506

Darren W. Henderson,
Center for Clinical and Translational Sciences, University of Kentucky, Lexington, Kentucky 40506

Keith W. Henry, and
Center for Clinical and Translational Sciences, University of Kentucky, Lexington, Kentucky 40506

Jeffery C. Talbert
Center for Clinical and Translational Sciences, University of Kentucky, Lexington, Kentucky 40506

Institute for Pharmaceutical Outcomes and Policy, College of Pharmacy, University of Kentucky, Lexington, Kentucky 40506

Abstract

We discuss informatics-based challenges of constructing large-scale collaborative networks for healthcare research and analysis from rural community health centers. These types of networks provide data access and analytic insights across multiple heterogeneous health centers for both healthcare professionals and biomedical researchers. Challenges fall into three general categories: data access, data integration, and technical infrastructure. Data access issues arise in balancing patient privacy, security, and utility; data integration issues persist from each site independently operating its desired electronic medical record; technical infrastructure challenges include creating an analysis and reporting hub capable of scaling across a large collaborative network. Other challenges, such as the difficulty of site recruitment, are important to discuss, but cannot be solved directly through informatics alone. We discuss these challenges and their potential solutions in the context of our implementation of the Kentucky Diabetes and Obesity Collaborative (KDOC). KDOC is a network of Federally-Qualified Community Health Centers (FQHCs) that established a collaborative infrastructure for research and analysis of obesity and diabetes in rural and under-served communities.

I. Introduction

Professional and organizational networks of health centers are long established with the purpose of overcoming the fragmentation of health service delivery in rural areas in terms of

---

coordinating regional services, health plans, recruitment, and training [1]. Despite these advances, a type of network established specifically for large-scale research and analysis of health data is relatively new. The relationship between professional networks, health collaboratives, and knowledge-sharing networks related to quality improvement leads to sustainable and well-functioning healthcare organizations, but not improved patient care [2]. An increased emphasis on research and analysis could potentially fill the gap between purely organizational networking and informatics-based networking to improve patient care. Evaluation methodology of clinical networks does exist [3], but not in the context of research and analysis.

Large-scale collaborative networks for research and analysis provide data access and analytic insights across multiple heterogeneous health centers. They establish a mutually beneficial relationship between the healthcare centers and network users, where the healthcare center gains crucial reporting abilities and analytic insights not provided by their local electronic medical record (EMR) and the network users potentially gain access to new and interesting data sets and possibly the ability to compare patient management and patient outcomes across all sites of the network.

Efforts such as i2b2 (integrating informatics from bench to bedside) [4] and the eMERGE (Electronic Medical Records and Genomics) network [5] have opened the door to quick cohort discovery in the clinical research domain. SHRINE (Shared Health Research Information Network) made it possible to federate i2b2 so that one query can target multiple healthcare sites [6]. Rural community health centers present their own challenges when it comes to data access and data integration, two of the key ideas central in establishing a useful shared network. Numerous previous studies documented challenges associated with adoption and implementation of EMRs in small rural practices and under-served communities [7], [8]. Without well-established EMRs, there is less chance for a potential network site to possess a subject-matter expert capable of stewarding the site’s healthcare data into the shared network. Even with a local expert, several informatics-based challenges exist in the context of data access, data integration, and technical infrastructure.

We discuss these challenges and their potential solutions in the context of our implementation of the Kentucky Diabetes and Obesity Collaborative (KDOC) [9]. KDOC is a network of Federally-Qualified Community Health Centers (FQHCs) that established a collaborative infrastructure for research and analysis of diabetes and obesity in rural and under-served communities. Figure 1 shows the initial phase of the KDOC network and its initial seven FQHCs from rural Appalachian Kentucky, a medically under-served area in which FQHCs and county health departments often act as the only safety-net providers available to low-income or uninsured patients.

II. Challenges

Challenges in building a collaborative network for research analysis fall into three general categories: data access, data integration, and technical infrastructure. Any of the issues that stem from these categories of challenges could dramatically reduce the impact and efficiency of the network; we will discuss each category in turn.

A. Data Access

Network topology can impact many facets of the network; generally speaking, research and analysis networks can be either distributed or centralized. In a distributed network each site is responsible for hosting data connected to the network, while in a centralized network the data from all sites is collected and made available at a central hub. For KDOC and many other networks that utilize rural community sites that do not have large information technology capacities, simply uploading their data to a central network hub is a far more realistic option than hosting a local data service.

Once a network is constructed, there are two types of data access challenges: connecting the healthcare site to the research and analysis network and connecting the network to researchers and healthcare professionals.

1) Site to Network Access—In our KDOC network, we observed three levels of access that each site maintains with its EMR: (a) access only through the EMR's graphical user interface (GUI), (b) GUI access and reporting tool access, (c) complete back-end database access. If the site only has GUI access to their EMR, they must contact their vendor to create a report that could feed into the network. Some sites might have the ability to create these reports themselves. If the site has complete back-end database access, they are free to extract information as necessary and create data feeds into the network. The obstacle with database access is that there may not be a subject-matter expert at the site capable of digging through a large stock of EMR tables to produce a report or feed for the network. Some sites may be willing to let a liaison from the network's organizers have access to either the reporting tools or even the back-end database access to assist in adding the site to the network.

2) Network to User Access—Once the network is established, access to the users of the network must be managed; management depends upon the specifics of the implementation chosen. For our KDOC project, we aggregated information from each site and produced dashboards in Tableau[1] that can act as an interface for interacting with the network's data; the details of the Tableau workbooks are given in Section II-C. To minimize development time, we created a master workbook and used parameterized access control to limit who has access to each site's data, as illustrated in Figure 2. In this user-level access scheme, a single workbook maintains a parameter indicating its site designation and is dynamic enough to show either each site's data, a selection of the network's data, or all of the network's data. The user connects to a workbook that excludes any data that is not allowed to be viewed (if any).

Each healthcare site will have concerns regarding patient privacy. To maintain patient privacy, we stipulated that each site be responsible for de-identifying their data before pushing their data to the network, making it impossible for network organizers to see private health information.

---

[1] http://www.tableau.com

B. Data Integration

It is highly unlikely that each site in the network uses the same EMR unless a common EMR was developed in advance [10] and without the luxury of a common EMR, data integration between sites becomes a challenge.

If every site is uploading or streaming data to the network, then great care and collaboration must take place to ensure the data is homogenized once it reaches the central hub; this requires a unified data model that every site must conform to when producing data. The data model can be as simple as files mimicking a dimensional model: a patients and demographics data file, a visits data file, a lab data file, a medications data file, and so on. Each dimension will need to agree on what controlled vocabulary or ontology must be used; successful networks like eMERGE hinge upon this idea and have reported the difficulties of the process [5]. It is highly likely that EMRs will mismatch on what vocabulary a dimension uses and a mapping between the local and the network choice must be made. In our KDOC project, we had positive results mapping internal laboratory codes to LOINC using RELMA [11], but verification was a manual process that required a human judge to assess correctness.

Data linkage between sites and even within a site can be problematic. It has been shown that infrastructure to support data linkages (disease registries, clinical data from EMRs, etc) could significantly impact translational research and patient outcomes, yet gaps in capacity across research and medical settings exist [12]. The biggest data linkage issue between sites is entity resolution: it is difficult to tell whether or not two patients are identical or not. Without a common EMR that maintains a master patient index, it is possible to over-estimate aggregate numbers when reporting across all sites. The link between entity resolution and information quality is well-established [13].

Within a healthcare site, entity resolution problems can still exist. If the fidelity between a patient’s visit and an external service such as laboratory orders or prescription medications is not maintained, it is difficult to link lab results or prescriptions to a specific visit of the patient’s history. The linkage between an observed laboratory result and a patient’s visit is necessary when searching or reporting on events that co-occur within a single visit. An example would be a report that lists patients that fall into an abnormal blood test range that did not have a diabetic screening on the same visit. In the worst case scenario where a linking visit identifier does not exist, entity resolution could potentially resolve observed facts such as a laboratory results with their true corresponding visit. Another example would be that some KDOC sites used multiple identifiers for the same physician if they were working at multiple internal locations, which made it difficult to report by physician.

Data and information quality is a major informatics issue [13], yet data quality remains difficult to quantify and data errors remains difficult to detect automatically. For our KDOC project, there were many cases of orphaned records that were not linkable by any other dimension. For example, it was common to see medical record numbers in the lab data that did not exist elsewhere. The quality of data was directly impacted by the quality of the technician working locally at the site and their knowledge of their internal EMR. If data elements are not picked from the most logical and fruitful locations, those elements may become irrelevant or distracting.

C. Technical Infrastructure

The decision to centralize the shared network was made with regards to the technical limitations of our rural health centers, as discussed in Section II-A. Centralizing the shared network's data feeds also necessitated centralizing the reporting hub used by healthcare professionals and biomedical researchers to interact with the shared data. We used Tableau to construct a series of reports and visualizations specific to KDOC's purpose of supporting obesity and diabetes research. Figure 2 illustrates how a master workbook can serve multiple sites and control access as discussed in Section II-A.2. Tableau's interface provides a point-and-click method to generate reports and visualizations. In Figure 3, our blood pressure measurement plot can quickly reveal patients who would benefit from better diabetes management. In Figure 4, rather than reporting information pertaining to the patient, the scorecard monitors how well the institution at large is monitoring diabetics. This enables benchmarking across all healthcare sites within the shared network. If an individual site is out-performing another in a particular measurement or category, that site can contribute to the shared network knowledge regarding protocols and/or policies to assist the struggling site in elevating scores in the problem area.

D. Site Recruitment and Other Challenges

Other challenges, such as the difficulty of site recruitment, are important to discuss, but cannot be solved directly through informatics alone. Site recruitment is the process of engaging a community health center and gaining their interest in participating in a collaborative network. The key to site recruitment is that network must provide a mutually beneficial relationship between the site and other sites or users of the network. It is well-established that EMRs often lack crucial capabilities needed for federal incentive and reporting programs [14]. The gap left from insufficiencies in the EMRs can be filled by joining a collaborative healthcare research and analysis network. Policy pressure and changing laws are linked to the closing of healthcare facilities [15]. The ability to effectively report and find crucial analytic insights cannot be undervalued and this must be effectively communicated to a potential site.

Even if a great case for joining the network is presented, other challenges do exist. Financial and geographic barriers to collaboration also occur. Regulatory issues are common-place when establishing business agreements between a new site and the shared network simply due to the sensitive nature of health data.

III. Conclusion

We reported our findings from implementing KDOC, a network of FQHCs serving rural Kentucky Appalachian areas, with the goal of providing infrastructure for diabetes and obesity research. Without networks like KDOC, valuable health data and outcomes would sit unused and disconnected from researchers. As a whole, these networks suffer from a lack of EMR standardization that prevents interoperable communication and inhibits analysis.

Additionally, each independent healthcare site has its own specific challenges when joining the research network that must be addressed through informatics-based principles and solutions.

Acknowledgments

Funding for this project was sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (RC4 DK089866) and the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR000117.

References

1. Moscovice I, Wellever A, Christianson J, Casey M, Yawn B, Hartley D. Understanding integrated rural health networks. Milbank Quarterly. 1997; 75(4):563–588. [PubMed: 9415092]
2. Cunningham FC, Ranmuthugala G, Plumb J, Georgiou A, Westbrook JI, Braithwaite J. Health professional networks as a vector for improving healthcare quality and safety: a systematic review. BMJ Quality & Safety. 2012; 21(3):239–249.
3. Carswell P, Manning B, Long J, Braithwaite J. Building clinical networks: a developmental evaluation framework. BMJ Quality & Safety. 2014; 23(5):422–427.
4. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association. 2010; 17(2):124–130. [PubMed: 20190053]
5. Pathak J, Wang J, Kashyap S, Basford M, Li R, Masys DR, Chute CG. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the emerge network experience. Journal of the American Medical Informatics Association. 2011; 18(4):376–386. [PubMed: 21597104]
6. Weber GM, Murphy SN, McMurry AJ, MacFadden D, Nigrin DJ, Churchill S, Kohane IS. The shared health research information network (shrine): a prototype federated query tool for clinical data repositories. Journal of the American Medical Informatics Association. 2009; 16(5):624–630. [PubMed: 19567788]
7. Bahensky JA, Jaana M, Ward MM. Health care information technology in rural america: electronic medical record adoption status in meeting the national agenda. The Journal of Rural Health. 2008; 24(2):101–105. [PubMed: 18397442]
8. King J, Furukawa MF, Buntin MB. Geographic variation in ambulatory electronic health record adoption: implications for underserved communities. Health Services Research. 2013; 48(6pt1): 2037–2059. [PubMed: 23800087]
9. Pearce KA, Jarrett T, Scutchfield FD, Bolt WD, Barron M, Houlihan JM, Dignan MB. Research partnerships with healthcare providers in rural community health centers, needs and challenges in diabetes research. Public Health Frontier. 2015; 4(1):1. [PubMed: 26457246]
10. DeVoe JE, Gold R, Spofford M, Chauvie S, Muench J, Turner A, Likumahuwa S, Nelson C. Developing a network of community health centers with a common electronic health record: description of the safety net west practice-based research network (snw-pbrn). The Journal of the American Board of Family Medicine. 2011; 24(5):597–604. [PubMed: 21900444]
11. McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, Forrey A, Mercer K, DeMoor G, Hook J, et al. Loinc, a universal standard for identifying laboratory observations: a 5-year update. Clinical Chemistry. 2003; 49(4):624–633. [PubMed: 12651816]
12. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: issues, methods, and directions for the future. Health Services Research. 2010; 45(5p2):1468–1488. [PubMed: 21054367]
13. Talburt, JR. Entity resolution and information quality. Elsevier; 2011.
14. Hsiao C-J, Decker SL, Hing E, Sisk JE. Most physicians were eligible for federal incentives in 2011, but few had ehr systems that met meaningful-use criteria. Health Affairs. 2012; 31(5):1100–1107. [PubMed: 22535501]

15. Ricketts TC. The changing nature of rural health care. Annual Review of Public Health. 2000; 21(1):639–657.

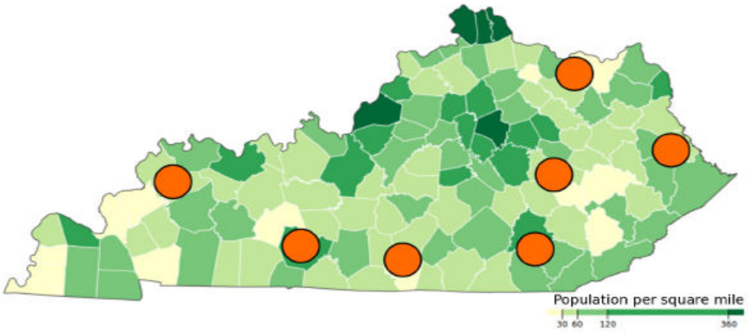

Fig. 1.

KDOC is a shared network of rural community health centers designed to support research and analysis of obesity and diabetes in Kentucky.

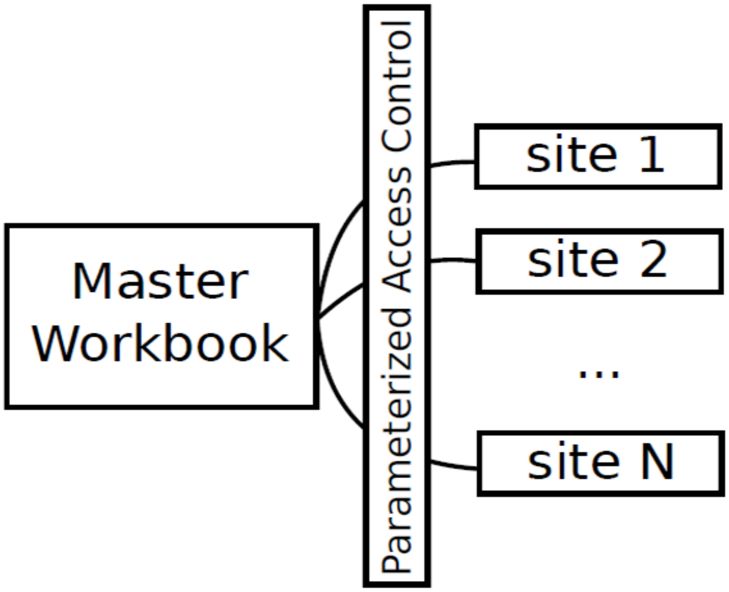

Fig. 2.

A parameter embedded into the workbook controls which sites are displayed in the data and corresponding visualizations.

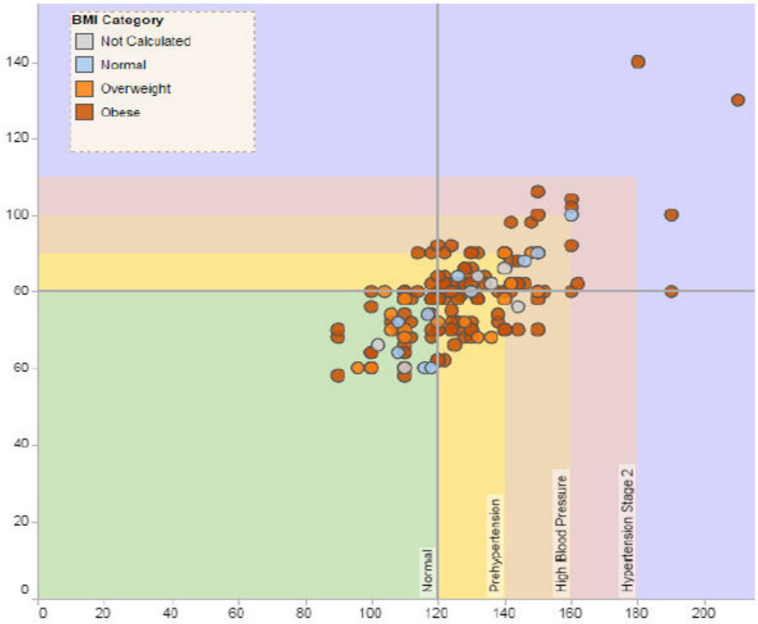

Fig. 3.

A simple plot of blood pressure (systolic vs diastolic) with the color of the circle indicating the patient's obesity level can quickly show outliers and patients that are in need of management.

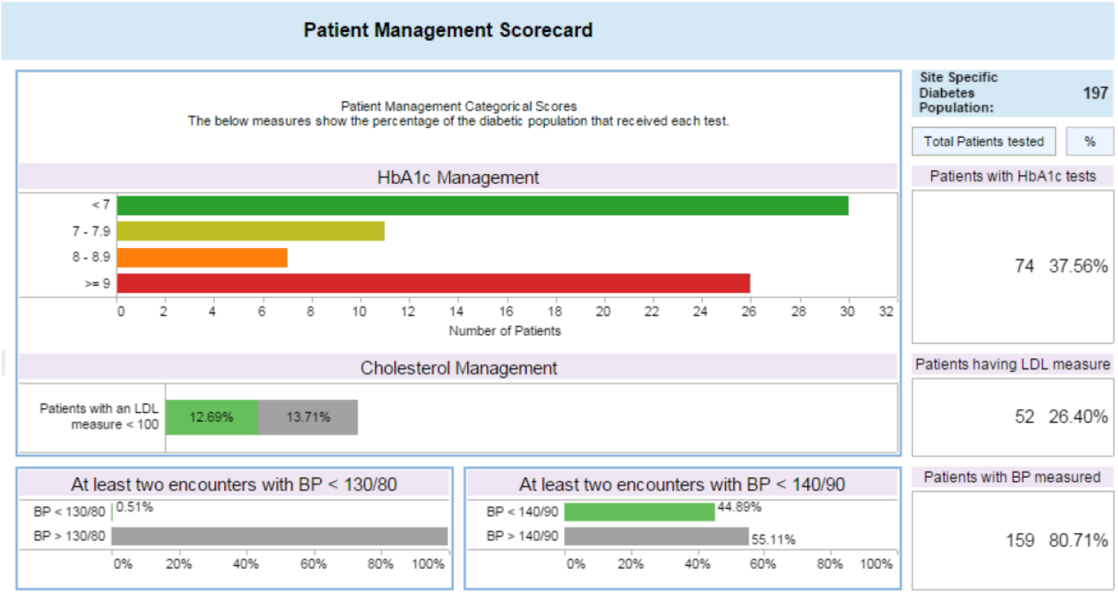

Fig. 4.

A patient management scorecard shows how well patients with diabetes are being managed by reporting if their HbA1c, cholesterol, and blood pressure is being monitored.